Supplemental Methods

Three cutting edge statistical concepts are being used here. The first is the notion of Object Oriented Data Analysis, see Marron and Alonso[1], which advocates treating each shape as a *data object*, essentially one data point or experimental unit. In the present context, this gives improved statistical power over usual approaches, which divide the information in the data among principal components, thus losing statistical power through multiple comparison summaries. The second is the use of the DiProPerm hypothesis testing algorithm of Wei et al[2]. This test is very useful in High Dimension Low Sample Size (HDLSS) contexts, such as the present. A main strength is that it maintains the potential for improved statistical power from the Object Oriented approach. Because it is a nonparametric permutation based test, assumptions are very mild, including only independence of cases, and commonality of class distributions. A tradeoff is that DiProPerm may have reduced power, relative to methods which rely on stronger assumptions. The third innovative analytical tool is the Distance Weighted Discrimination (DWD) machine learning algorithm[3]. This is presently used to find a good separating direction between classes, both in the visualizations and in DiProPerm. Again assumptions are very minimal, because DWD is not based upon any probability distributions. Its strength is that it gives better than usual separation of projected classes, especially in HDLSS contexts.

The 4 steps used in our approach are elaborated below.

> *1. Generate a visual display of differences between distributions using scatter plot matrices obtained from principal component (PC) directions as well as from the distance weighted discrimination (DWD) direction.* Scatter plots were based on 2 dimensions out of 120 dimensions of data. PCA finds an orthonormal basis which guarantees the maximum variances of the projections of the data. Hence, we use principal component directions as the x- and y-axis of scatter plots since better visualization may usually be obtained from greater dispersion of the data.

Next, we utilized the DWD vector as a projection direction. DWD is a machine learning-based linear discriminant analysis method which maximizes the distances between a separating hyperplane and the data points [1]. DWD was developed to address the shortcomings of Support Vector Machine (SVM), and has been shown to be at least as good as other methods in finding a projection direction on which two groups are well discriminated. DWD has particularly good performance characteristics when the dimension of a data vector exceeds the number of samples, namely in high dimensionality low sample size (HDLSS) settings.

2. Perform naive 2-sample t-tests on each set of projection scores with Bonferroni correction for multiple comparisons to determine which differences are statistically significant. A statistical test is needed to confirm a potential separation seen in Step 1. The Hotelling T^2 statistic using all 120 dimensions might be considered to test the hypothesis that the group means are the same. However, in a HDLSS setting where the sample size is less than the dimension (e.g., the number of hips from African American females [n=49] is less than their dimension [n=120]), the Hotelling T^2 statistic is not applicable. As a naive method for testing whether the mean group differences are equal, we first project the data into a subspace spanned by the first 4 principal directions, and then apply a 2-sample t-test to each set of PC scores. Although we cannot obtain test results on statistical differences in the original 120-dimensional space, we can regard the individual t-test results as an indicator of the existence of any significant differences.

3. Perform the potentially more powerful Distance Projection Permutation (DiProPerm) test [2] to determine statistically significant differences between the two distributions in a fully multivariate fashion. Drawbacks of the t-tests based on PC scores used in Step 2 include a potential loss of information (i.e., from only considering the first several components), and a lack of focus on information in the data (i.e., most components are not expected to contribute in a useful way). These issues are directly remedied by the DiProPerm test for the difference between two distributions [2] which was specifically designed for HDLSS situations. Given data with known groups (e.g., case and control) and a procedure for obtaining a direction on which projected data of the groups are well separated (e.g. DWD), the DiProPerm procedure performs a hypothesis test following 3 steps:

1. Direction : Obtain the direction of projection using DWD.
2. Projection : Project the data onto the direction from Step 1 and obtain the mean difference [MD], defined as the difference between the sample means of the projections of two classes on a projection direction. Using the DWD direction as a projection direction and the MD statistic as a test statistic yields maximum power among combinations of various projection directions and test statistics[2].
3. Permutation : Derive the p-value of the obtained statistic through use of a permutation test. That is,
    i. Permute the original group labels.
    ii. Derive the projection direction from the permuted class labels.
    iii. Project the data onto the direction in (ii) and obtain a simulated statistic.

Repeat the steps (i-iii) a specified number of times (e.g., 1000), so that a sample of simulated statistics is obtained. The p-value of the original statistic is computed as the ratio of the number of the simulated statistics larger than the original statistic to the total number of the simulated statistics.

The capability of the DiProPerm test to handle the HDLSS setting is critical in our analysis since it enables us to perform analyses despite relatively small sample sizes. For example, although the dimension of each femur is p=120, the DiProPerm test still enables us to draw statistical conclusions on the femur shape of African American females (n=49).

4. Plot the femur shape variation to give anatomical insights into the direction and amount of variation between the two groups.

This 4-step analysis also allows us to compare the strength of these associations, e.g., we see a stronger association of femur shape with incident RHOA case status among men than among women.

Sensitivity analyses:

We have used a more powerful statistical method for our analyses, but it is also a new method and there is not yet a clear manner by which to adjust for covariates. The DWD method uses classification, and to include age, BMI, etc. would require numerous classes and would thus reduce sample size in the individual groups. Ideally, in order to "adjust" the DWD method described here, new statistical methodology would have to be developed. However, we can convert the DWD vector into a score and use that as an outcome in a standard regression model to allow for adjustment for age, BMI, and baseline KL grade which we have added as a sensitivity analysis with additional detail below:

To adjust for the effects of age, BMI and baseline KL grade in our analysis, we fit an ANalysis of COVAriance (ANCOVA) model to the DWD score. Using obvious notation, for the i-th patient, we regress the DWD score on KL grade (0 or 1), age and BMI. In other words we fit a linear model

$$(\text{DWD Score})_\text{i} = \mu + \alpha(\text{KL grade})_\text{i} + \beta_1(\text{age})_\text{i} + \beta_2(\text{BMI})_\text{i} + \epsilon_i$$

where μ is the overall mean of the DWD scores, α is the relative effect of 1 KL grade to 0 KL grade, β_1 and β_2 are the regression coefficients of age and BMI respectively and ϵ_i is the random error of the i-th patient. By fitting the model to the DWD scores using least squares estimation, we can obtain the estimates for the parameters $\hat{\mu}, \hat{\alpha}, \hat{\beta}_1$ and $\hat{\beta}_2$. Now we define the adjusted DWD score defined as

$$(\text{Adjusted DWD Score})_i = (\text{DWD Score})_i - \hat{\mu} - \hat{\alpha}(\text{KL grade})_\text{i} - \hat{\beta}_1(\text{age})_\text{i} - \hat{\beta}_2(\text{BMI})_\text{i}.$$

Using the adjusted DWD scores, we perform the same analysis for the difference between cases and controls for African American women.

As above, methods for adjusting DWD do not currently exist, and there is no equivalent to Generalized Estimating Equations (GEE) or clustering to apply here to account for potential non-independence between hips. However, as an additional sensitivity analysis, we have repeated the DiProPerm tests, randomly removing one hip from each person who had two hips in the dataset (n=40).

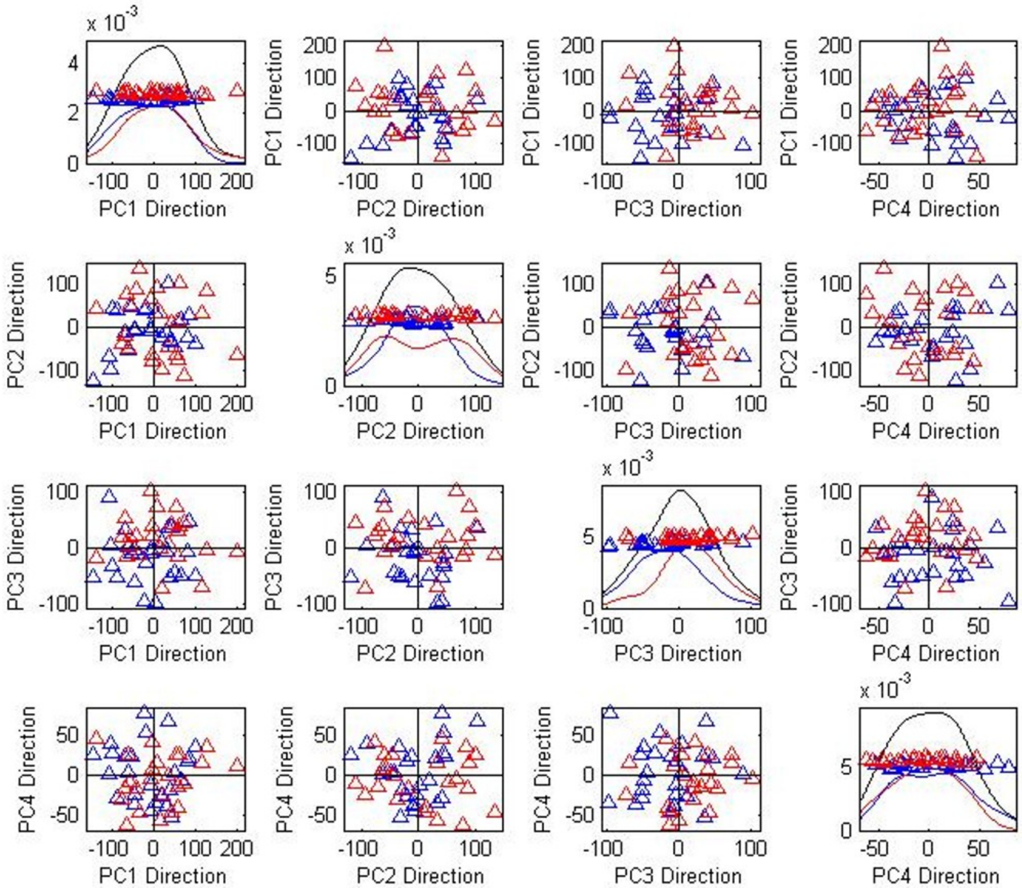
After removal, the number of observations (hips) decreased from 382 to 342 and the number of hips from African American women decreased from 49 to 45. The DiProPerm test for the difference between cases and controls among African American women now yields a p-value of 0.063. Since we are removing random side of each hip of the 40 participants, the test results vary depending on which hip we include in our analysis. 5 runs of random selection yields the following 5 p-values:

p-value	0.0642	0.0578	0.0700	0.0366	0.0629

One might wonder what result will be obtained if we adjust for the confounding factors as mentioned above as well as perform the DiProPerm test on the same data set used here (excluding bilateral hips). This adjustment decreases the p-value to 0.036, which is below the significance level and indicates a significant difference between cases and controls after adjustment for age, BMI and baseline KLG, and is consistent with our initially reported unadjusted results.

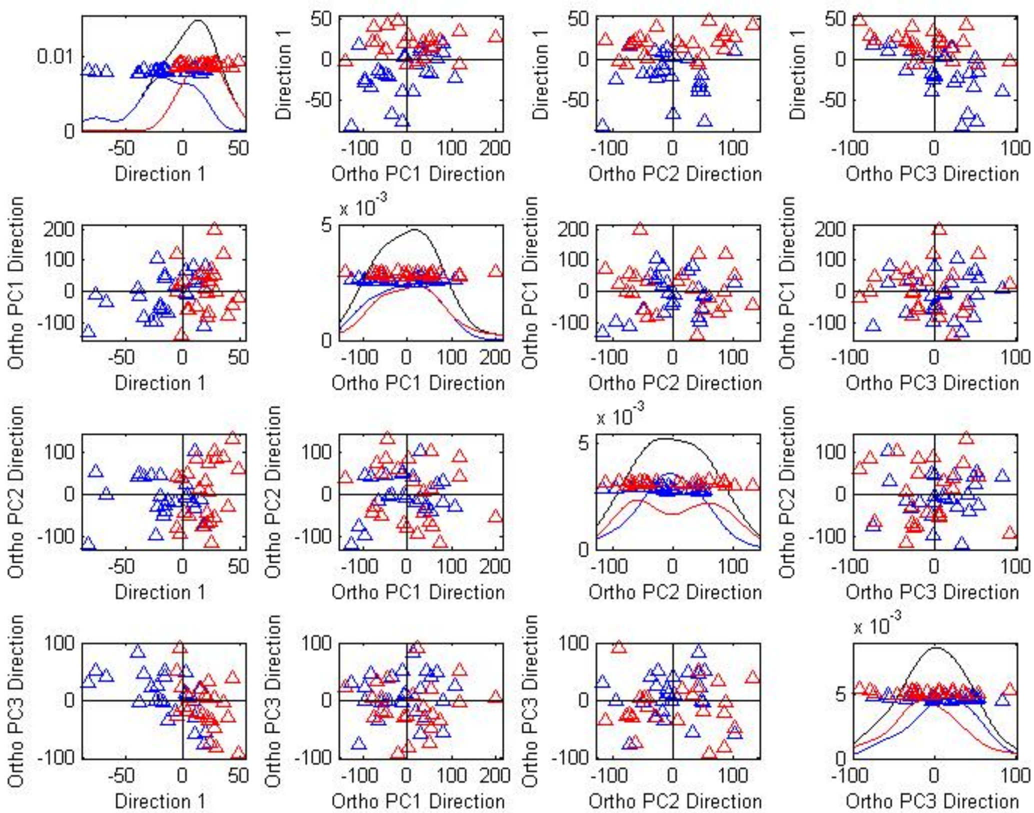
References

1. Marron JS, Todd MJ, Ahn J: **Distance-weighted discrimination**. *Journal of the American Statistical Association* 2007, **102**(480):1267-1271.
2. Wei S, Lee C, Wichers L, Li G, Marron JS: **Direction-projection-permutation for high dimensional hypothesis tests**. In: *arXiv.* 2013.

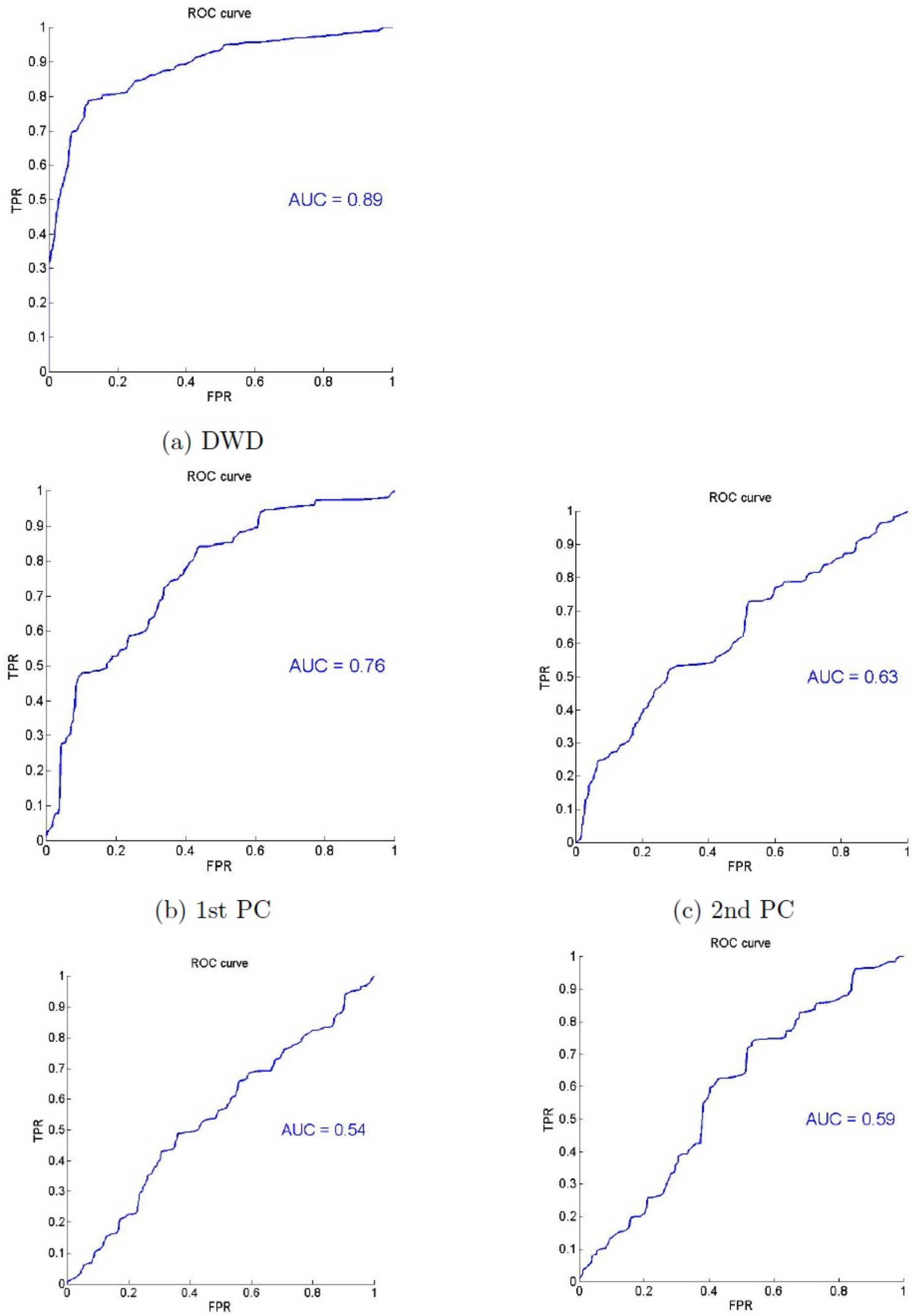


Score	PC 1	PC 2	PC 3	PC 4
p-value	0.886	1.000	0.052	1.000

Supplemental Figure A: Scatter plots of the PCs of African American women (traditional method). Incident RHOA cases are red, and controls are blue. The top left graph shows a distribution of the projections of the data onto the 1st principal direction, where the black curve represents the smoothed histogram of all the 1st PC scores, while the red and blue curves show the smoothed histograms of the scores of the incident RHOA cases and controls, respectively. The table shows Bonferroni adjusted p-values from 2-sample t-tests on PC scores for the difference between African American women who are incident RHOA cases and those who are controls.



Supplemental Figure B: Scatter plots of the projections of African American females on the DWD (novel method) and orthogonal principal directions. Incident RHOA cases are red, and controls are blue. In this figure, the first orthogonal principal direction is such that the projections of the data onto the direction have the maximum variance among all projections on the directions orthogonal to the DWD direction. Similarly, the second orthogonal principal direction is the direction that guarantees the maximum variance of the projections among all directions orthogonal to the DWD and first orthogonal principal direction. The DWD direction (upper left) gives better visual separation of cases and controls than any PC direction seen in Supplemental Figure A.



(a) DWD

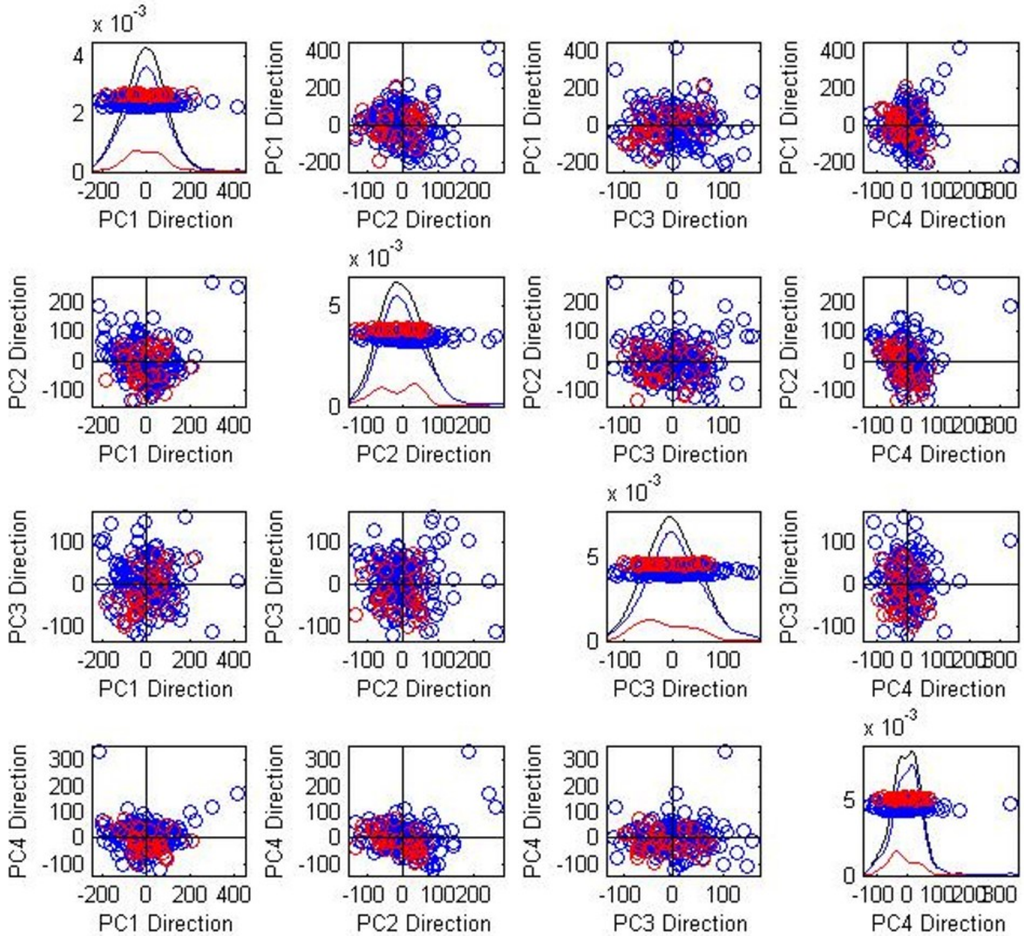
(b) 1st PC

(c) 2nd PC

(d) 3rd PC

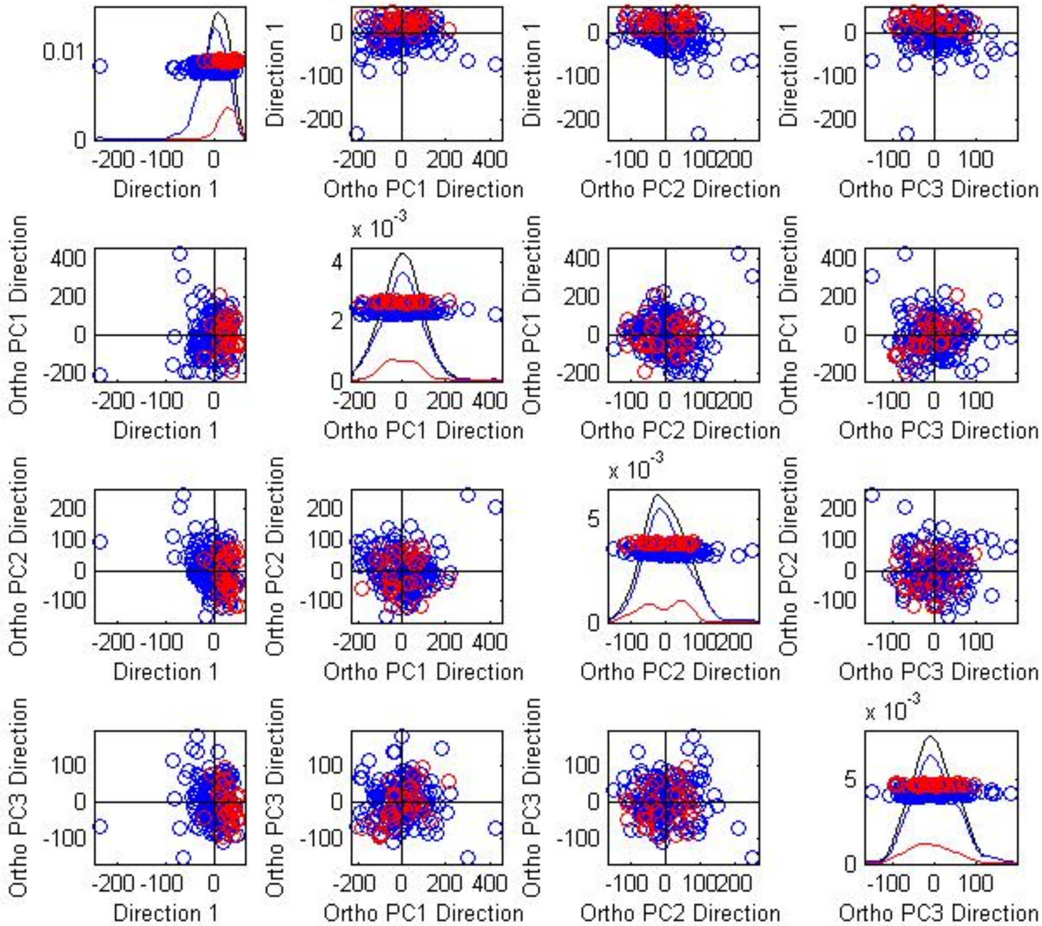
(e) 4th PC

Supplemental Figure C: Receiver Operating Characteristic (ROC) Curves for sex among incident RHOA cases. Panel (a) shows the ROC curve and AUC for the discrimination rule for the sex effect on incident RHOA cases based on the projections on the DWD direction, while (b), (c), (d) and (e) show the corresponding ROC curves for the discrimination rule for sex on incident RHOA cases based on the projections on the 1st, 2nd, 3rd and 4th PC directions, respectively.

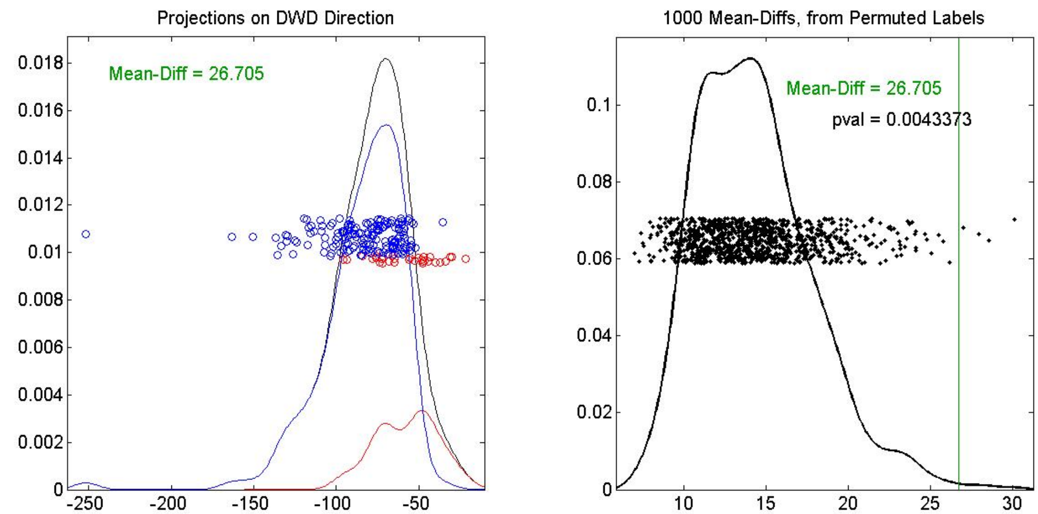


Score	PC 1	PC 2	PC 3	PC 4
p-value	1.000	0.496	0.167	0.346

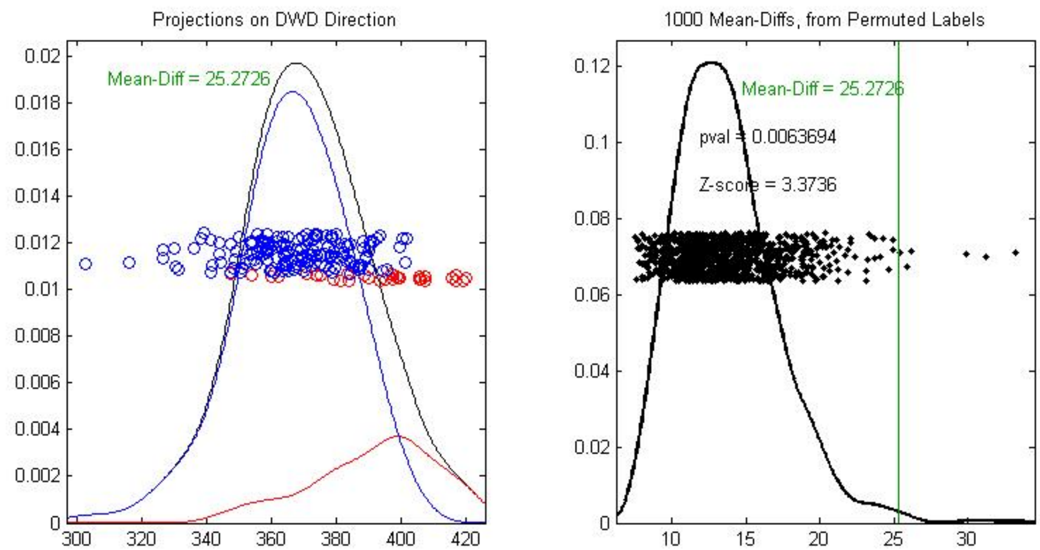
Supplemental Figure D: Scatter plots of principal scores of incident RHOA cases by race (traditional method). African Americans are red, and whites are blue. Bonferroni-adjusted p-values are given for the first 4 principal components' scores' 2-sample t-tests.



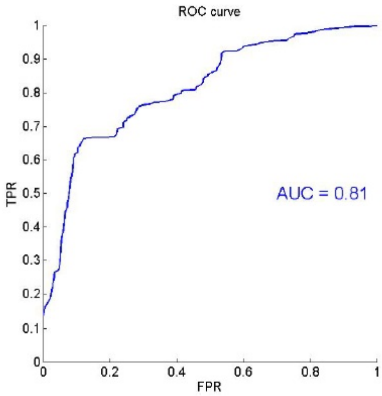
Supplemental Figure E: Scatter plots of the projections of incident RHOA cases by race on the DWD and orthogonal principal directions (novel method). African American are red, and whites are blue.



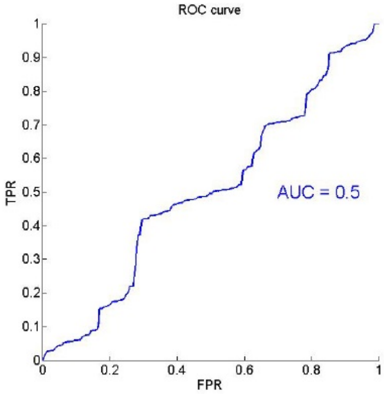
Supplemental Figure F: DiProPerm test result for the difference in the proximal femoral shape between African Americans and whites, among cases of incident RHOA. Blue represents the whites and red represents the African Americans.



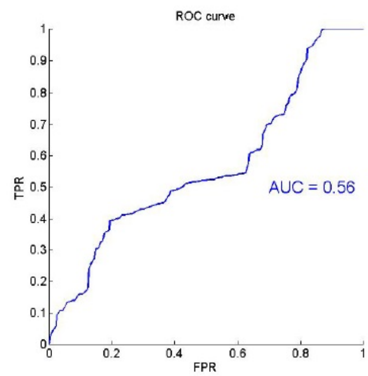
Supplemental Figure G: DiProPerm test result for the difference in the proximal femoral shape between African Americans and whites, among controls who did not develop incident RHOA. Blue represents the whites and red represents African Americans.



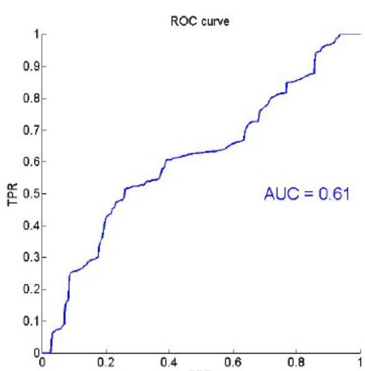
(a) DWD



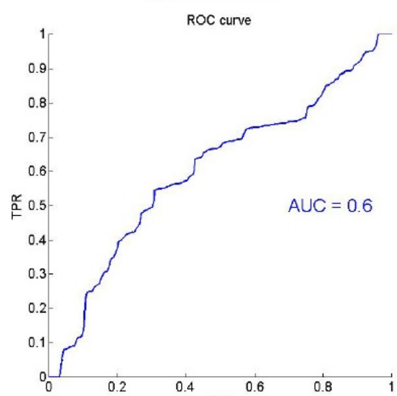
(b) 1st PC



(c) 2nd PC



(d) 3rd PC



(e) 4th PC

Supplemental Figure H: Receiver Operating Characteristic (ROC) Curves for race among incident RHOA cases. Panel (a) shows the ROC curve and AUC for the discrimination rule for the race effect on incident RHOA cases based on the projections on the DWD direction, while panels (b), (c), (d) and (e) show the corresponding ROC curves for the discrimination rule for sex on incident RHOA cases based on the projections on the 1st, 2nd, 3rd and 4th PC directions, respectively.